International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

Smart Automated Modelling using Eclat Algorithm for Traffic Accident Prediction

Deekshitha H R[1], K R Sumana[2], Dr H D Phaneendra[3]

[1]*PG Student, Department of MCA, NIE, Mysuru, India*
[2]*Assistant Professor, CS Division, NIE, Mysuru, India*
[3]*Professor & Head, CS Department, NIE, Mysuru, India*

Abstract - *The number of traffic accidents are rapidly increasing which leads to increase in deaths and injuries. Road safety becomes an important key factor to be concentrated to reduce all of these accidents. In this paper, we have concentrated on the study of the factors that are influencing the road accidents and the accident patterns that are frequently occurring. The main aim here is to identify those hidden rules. The hidden rules in the frequently occurring accident item set uncovers the association between the factors that are influencing the accidents. The occurrence of accidents can be reduced by breaking these hidden rules. Identifying these patterns helps in taking precautionary measures towards road safety. Data Mining is an emerging technology that can be used to mine the frequent item set and patterns based on the association rules like support count and confidence. Association rule mining technique called "Eclat algorithm" used in this paper to find the frequent itemset and determine the accident patterns that are frequently occurring. This algorithm is effective and faster than "Apriori algorithm" because the dataset will be fetched only once and the process is carried out by refering to the previous tables. By developing the system and obtaining the patterns it was seen that this technique is effectively working with good results. Thus, a smat automated modelling using ecalt algorithm for traffic accident prediction was successfully developed that helps the traffic department to concentrate on road safety by avoiding frequently occurring accident patterns.*

Key Words: **Data Mining, Eclat Algorithm, Apriori Algorithm, Association Rules, Road Safety.**

1. INTRODUCTION

Nowadays number of traffic accidents are rapidly increasing which leads to increase in deaths and injuries. Road safety becomes an important key factor to be concentrated to reduce all of these accidents and reduce deaths and injuries. Discovering the associations among the traffic accidents is the key factor in reducing the traffic accidents. Data mining is an emerging technology which process data and extracts useful information from the processed data.

Smart automated modelling using eclat algorithm for traffic accident prediction is a paper that tells us about the analysis of the traffic accidents data, to discover the useful traffic accidents patterns. Here we make use of data mining technology to analyse the data. Predictive mining applied on past data about already occurred accidents in union with any other key information weather the traffic condition creates an compelling alternative with likely meaningful and beneficial results for all involved participants. Precautionary measures can be taken to reduce the traffic accidents based on these analysis.

The system that we are proposing is a real time application, in which the hidden strong rules in the frequent itemsets will be discovered to reveal the association between the factors that are influencing the accidents. These hidden rules may be useful to break the similar accident patterns which in-turn results in reducing the number of accidents. This extracted knowledge will be help to improve the road safety that results in reducing deaths and injuries. Here the system uses old traffic accidents data to extract frequent patterns and key factors causing different accidents types. Proposed system depicts the associations among traffic accidents by making use of "*Association Rules*" to find the accident patterns.

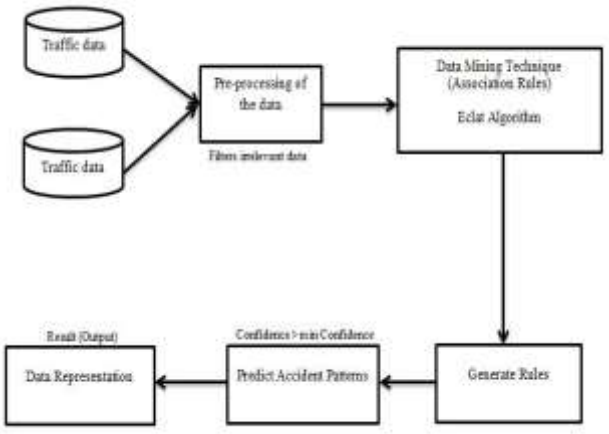

Fig -1: System Architecture

In the fig 1 shown above the complete system architecture is depicted. The first step is to fetch the data from the database. The traffic data will be pre-processed and the irrelevant data will be erased. Then that processed data will be given as the input to the main module of the system called eclat algorithm. Eclat algorithm is an association rule mining technique used to extract the frequent patterns. Then the

rules will be generated to classify those patterns. After that using association rules certain strong frequent pattern are generated. That patterns are given as the output of the system.

2. LITERATURE SURVEY

The use of appropriate analytical methods to improve road safety is an interesting challenge, especially in the current situation, when the amount of data available increases continuously. In this case, it is possible to select a series of existing studies mainly focused on local road accident data. Iran is one of the countries with a high mortality rate due to road accidents and injuries. In the last three years, road accidents in Iran have caused 24,000 people (i.e 3 people per hour) in the average number of deaths and about 240,000 injured per year [2], [3]. This fact has motivated a group of authors to identify the most important factors that influence the severity of the injuries of drivers involved in road accidents on these roads [3]. They used incident data from the records of the Department of Information Technology of Iran's Traffic Police from 2006 to 2008. Each incident was described by 14 attributes. The goal was the severity of the wound with 3 levels: no wounds, wounds and fatalities. The records include over 169,000 drivers. The selection of the factors was made on the basis of the Measure of Variable Importance (VIM), which is one of the results of the CART (Classification and Regression Tree) method. The results indicated that the seat belt was the most important factor associated with the severity of the injury from road accidents, and non-use significantly increases the likelihood of being injured or killed.

Authors Chang and Wang [4] used a similar approach to generate the CART model to identify the relationships between the severity of the injury and the characteristics of the driver / vehicle, the road / environmental variables and the impact variables. Using data from the 2001 crash for Taipei (Taiwan), they identified the vehicle type as the most important variable for the severity of the crash. Furthermore, data from central Taiwan were used in the study by authors Yau-Ren Shiau et al. [5] to identify at the beginning the most important factors that affected more than 2,400 road accidents in 2011 and, subsequently, extraction methods such as Fuzzy Robust Main Component Analysis, Backpropagation Neural Network and Logistic Regression were applied to generate classification models expected The best accuracy of 85.89% was obtained by combining the first two methods mentioned.

In Dubai, a study was dedicated to road safety. About 600 people die every year in road accidents, in United Arab Emirates. Traffic accidents are the second leading cause of death [6]. From 2008 to 2010, the authors used over 1,800,000 records with 19 attributes covering accident, driver and road / vehicle conditions. The objective attribute (the degree of seriousness of the accident) contained five values: fatal, serious, moderate, minor and none. WEKA support tools were used to create the final predictive models and the following algorithms: J48 decision trees, Bayesian networks, PARTS and multilayer Perceptron. Multistrato Perceptron generated the most accurate model (over 99%) and the Bayesian network gave maximum speed to build the model (0.17 seconds).

All the studies presented have different characteristics in common: they used a limited sample of data from a specific period of time or place; it may be considered positive to take local characteristics into consideration, but, on the other hand, some important global factors may be omitted; In many cases, traditional data mining algorithms such as decision trees have apparently been used due to their simple comprehension and interpretation. Virtually all studies have led to models with a high percentage of accuracy, but this value largely depended on the methods of selecting the features used. The experiments described in the project concerned a relatively large volume of data that requires new methods and support environments.

3. TECHNIQUES INVOLVED

3.1 Data Mining

The group of facts and observations are nothing but the data. The process of analysing those data from different prospect and summarizing that data into a benificial information is called data mining. It allows us to analyse the data from distinct dimentions and then categorize it to identify the relationships.

3.2 Association Rules

In this project work, we have used a data mining technology called association rules to find the correlations in the transactional data. It uses *if-then* associations/patterns to find out the most important key relationships in the provided dataset by making use of *support* and *confidence* criteria. The indication of how frequently the items appear in the transactions is called support. Confidence indicates the number of items the if-then statements are found true.

3.3 Eclat Algorithm

We make use of a data mining algorithm called *eclat* (Equivalent Class Clustering and Bottom Up Lattice Traversal) algorithm. This algorithm is used for frequent item set generation. It searches the data in a depth first search(DFS) pattern/manner. The data is represented in a vetical format.

The basics of eclat algorithm is taken from another association mining algorithm called *apriori algorithm.* The main disadvantage of apriori algorithm is that whenever it needs to calculate the support it reaches the database and take out the data from it, which makes the system work slow. But in eclat algorithm once the transactional data and the item set is fetched from the database it only refers to the

previous table to calculate the support. This makes the system work much faster and efficiently than the apriori algorithm.

4. ECLAT ALGORITHM CONCEPT

Association is probably the better known and most familiar and straightforward data mining technique. Here, a simple correlation is made between two or more items of the same type to identify patterns. In our project work, association rule mining "Eclat Algorithm" is used to predict the relationship between different traffic accidents using the traffic dataset.

Algorithm 1: Working of Eclat Algorithm

Step 1: Get T.Id list for each item (Database or Any File)

Step 2: T.Id list of {a} is exactly the list of transactions containing {a}

Step 3: Intersect T.Id list of {a} with the T.Id lists of all other items, resulting in T.Id lists of {a,b}, {a,c}, {a,d}, ...= {a} - conditional database (if {a} removed)

Step 4: Repeat from 1 on {a} - conditional database

Step 5: Repeat for all other items

In this algorithm the item sets are identified from different transactional data. Once the dataset are taken from the database it stores it in a table format. For each item the support is identified, (i.e In how many transactions that particular item exists). Once the support is identified for every item in the item set it is saved in table called candidate itemset C1. the next step is to generate the frequent item set table L1. Frequent itemset table L1 contains items that are greater than the minimum support count, (i.e If the minimum support count is given as two then the items that exists in lesser than two transactions are eliminated). Once the frequent item set table L1 is ready then keeping that as a base table, the items are combined with every possibility of each item, again support is identified for those combined items. That data is stored in table called candidate item set C2. Then again for that frequent item set table L2 is calculated. It keeps on repeating until the items cannot form a subset with the item set. Atlast all the data from frequent item set like L1,L2,L3… are taken together as frequent item set (L). From the frequent item set table confidence is calculated for the items that have some relationships, (i.e No of transactions containing each of the item/ No of transactions containing only one item). Then finally strong association rules are generated, (i.e the relationship confidence between two items that is greater than the specified minimum confidence).

The items that possess greater confidence are displayed as the identified patterns from the past data. That pattern will be diaplayed to the user which was analysed from the past data collected and stored in the database.

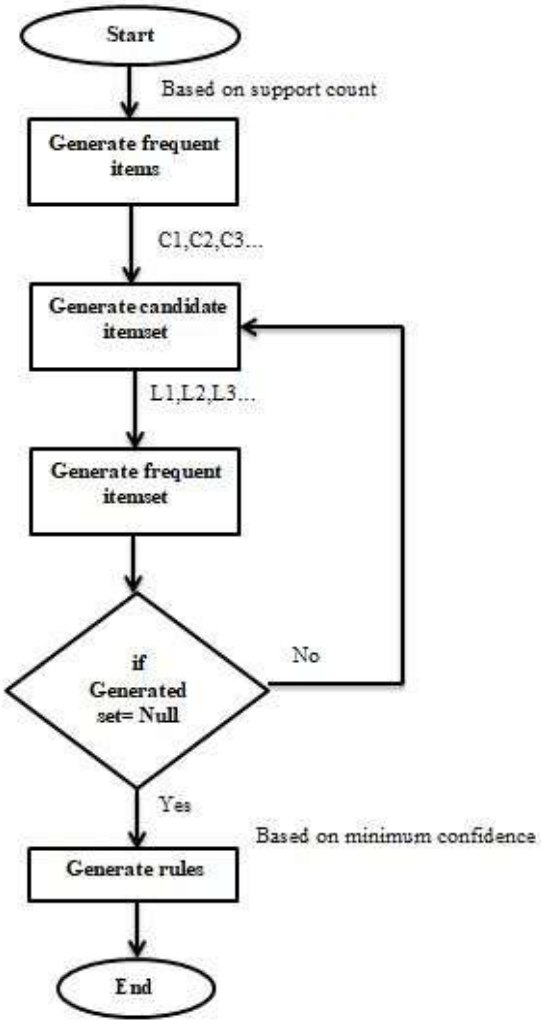

Fig -2: Flowchart of Eclat Algorithm

5. METHODOLGY FOR DEVELOPING THE SYSTEM

In this paper we are briefely explaining how to develop a system that predicts patterns based on the frequent item set. For this work we are using Eclat algorithm as the main basis. The table -1 shown below gives the format of how the accident datasets are stored in the database. Based on these dataset the user will be able to see the patterns of accidents that are occuring frequently. Accident type may include Over Speed, Single Car Collision, Two Car Collision, Bike Collision, Hit and Run, Drunk and Drive/Medical Drugs, Sporting Accident, Fatal Accident or any other type of accidents.

Table -1: Dataset Sample

Road Id	Description	Occurred Date	Accident Type
01	Death Major Injuries Small Injuries	10-09-2016 10-09-2016 10-09-2016	Over Speed Two Car Collision Bike Collision
02	Lack of attention Skid Death	02-03-2015 02-03-2015 02-03-2015	Hit and Run Over Speed Drunk and Drive/Medical Drugs
03	Hitting Barrier Driver's Fault Skid	01-01-2014 01-01-2014 01-01-2014	Single Car Collision Drunk and Drive/Medical Drugs Sporting Accident
04	Due to Rainfall Inexperience Death	06-02-2015 06-02-2015 06-02-2015	Fatal Accident Sporting Accident Hit and Run

Algorithm 2: Automated Modeling for Traffic Accident Prediction

Step 1: Fetch the data from the database based on the interested item set.

Step 2: Declare the minimum support count and minimum confidence.

Step 3: The fetched data will be given as input to the eclat algorithm which gives the frequent item set as the output based on the given support and confidence.

Step 4: Based on the minimum confidence declared the item patterns which has stronger confidence will be displayed.

Depending on the dataset fetched from the database the process of prediction begins. Here the minimum support count and minimum confidence must be declared at first. Then data will be classified. The transactional id (T.ID) is nothing but accident occurred dates in this system. Those data will be fed to the eclat algorithm, as explained before in *Algorithm 1,* the frequent item set will be generated from minimum support count and minimum confidence specified. The patterns which has greater confidence than specified confidence, those are considered as frequently occurring traffic accident patterns

6. CONCLUSIONS

In this paper we have explained a system that can be developed which predicts patterns of accidents automatically that may occur in the roads in future days, so that the government can take precautionary measures towards road safety. This system can be developed for traffic departments who manages the roads and road safety. These predicted patterns can be useful for the police who can take safety measures at the right time that helps in reducing the accidents which in-turn reduces the injuries and deaths.

In future a graphical view for the prediction shown can also be developed for the system, so that it will be more user friendly.

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my guides K R Sumana and H D Phaneendra for their valuable guidance throughout. Secondly, I would also like to thank my parents for their blessings and colleagues for their help.

REFERENCES

[1] Zhen Gao et al, " Research on Automated Modeling Algorithm Using Association Rules for Traffic Accidents", IEEE International Conference on Big Data and Computing, 15-17 Jan 2018, Electronic ISSN- 2375-9356

[2] F.M.O.I. Forensic Medicine Organization of Iran; Statistical Data, Accidents, online avail-able on: http://www.lmo.ir/?siteid=1&pageid=1347.

[3] A.T. Kashani et al., "A Data Mining Approach to Identify Key Factors of Traffic Injury Severity", PROMETTraffic& Transportation, 23(1), pp. 11-17, 2011.

[4] L.Y. Chang, H.W. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques", Accident Analysis and Prevention, 38(5), pp. 1019-1027, 2006.

[5] S. Yau-Ren et al. "The Application of Data Mining Technology to Build a Forecasting Model for Classification of Road Traffic Accidents", Mathematical Problems in Engineering, Volume 2015 (2015), pp. 1-8., 2015. F. Babi and K. Zuská ová • Descriptive and Predictive Mining on Road Accidents Data– 92

[6] A Araar et al., "Mining road traffic accident data to improve safety in Dubai", Journal of Theoretical and Applied Information Technology, 47(3), pp. 911-927, 2013.